EXHIBIT A

Introduction to Protein Structure

Second Edition

Carl Branden

Microbiology and Tumor Biology Center
Karolinska Institute
Stockholm
Sweden

John Tooze

Imperial Cancer Research Fund Laboratories
Lincolns Inn Fields
London
UK

THE COVER

Front: The structure of the potassium channel from *Streptomyces lividans*, determined by Rodney MacKinnon at the Rockefeller University, New York. As discussed in Chapter 12, this structure—the first of such an ion channel—shows how the channel allows the passage of potassium ions through cell membranes with high efficiency and selectivity. The view is looking down the protein as it sits in the cell membrane, as seen from outside the cell, with a potassium ion shown in gold. This image was produced using the GRASP program (A. Nicholls and B. Honig, Columbia University) from atomic coordinates kindly provided by Rodney MacKinnon.
Back: A hand-drawn image of the potassium channel, in the same view as on the front cover, with each subunit of the tetrameric protein shown in a different color.
Cover design by Christopher Thorpe and Nigel Orme.

Visit the *Introduction to Protein Structure* Web site:
http://www.proteinstructure.com/

For information on other textbooks available from Garland Publishing, visit:
http://www.garlandpub.com/

Library of Congress Cataloging-in-Publication Data
Branden, Carl-Ivar
Introduction to protein structure / Carl-Ivar Branden, John Tooze.
--2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-2304-2 (hardcover). --ISBN 0-8153-2305-0 (pbk.)
1. Proteins--Structure. I. Tooze, John. II. Title.
QP551.B7635 1998
572'.633--dc21 98-34487
CIP

Published by Garland Publishing, Inc.
19 Union Square West, New York, NY 10003-3382

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

The Building Blocks

1

Recombinant DNA techniques have provided tools for the rapid determination of DNA sequences and, by inference, the amino acid sequences of proteins from structural genes. The number of such sequences is now increasing almost exponentially, but by themselves these sequences tell little more about the biology of the system than a New York City telephone directory tells about the function and marvels of that city.

The proteins we observe in nature have evolved, through selective pressure, to perform specific functions. The functional properties of proteins depend upon their three-dimensional structures. The three-dimensional structure arises because particular sequences of amino acids in polypeptide chains fold to generate, from linear chains, compact domains with specific three-dimensional structures (Figure 1.1). The folded domains can serve as modules for building up large assemblies such as virus particles or muscle fibers, or they can provide specific catalytic or binding sites, as found in enzymes or proteins that carry oxygen or that regulate the function of DNA.

To understand the biological function of proteins we would therefore like to be able to deduce or predict the three-dimensional structure from the amino acid sequence. This we cannot do. In spite of considerable efforts over the past 25 years, this folding problem is still unsolved and remains one of the most basic intellectual challenges in molecular biology.

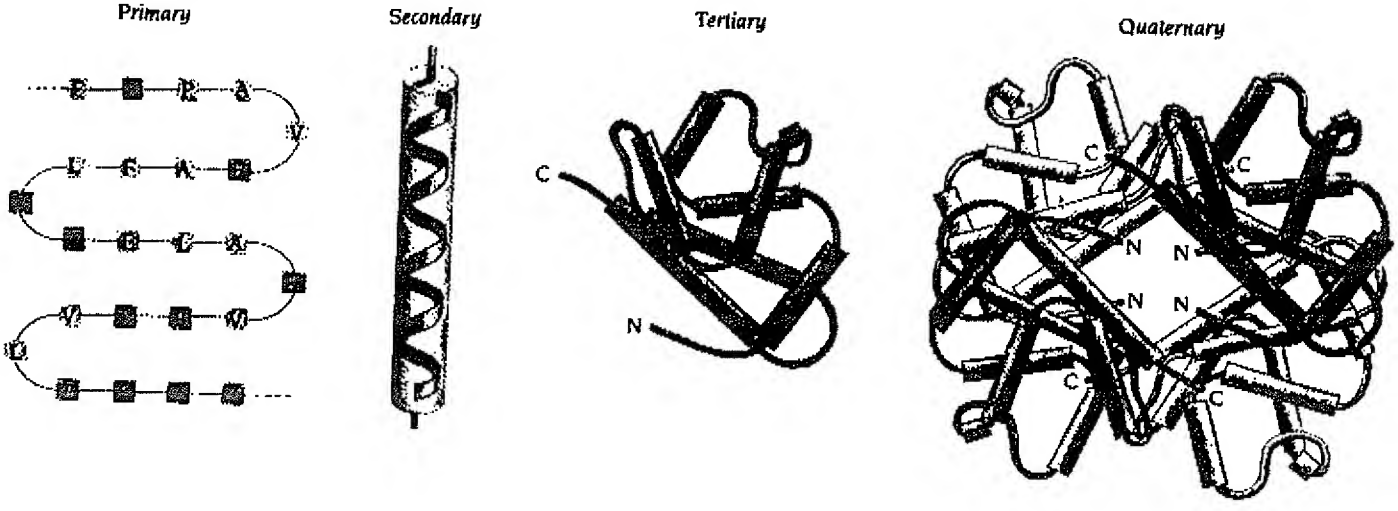


Figure 1.1 The amino acid sequence of a protein's polypeptide chain is called its primary structure. Different regions of the sequence form local regular secondary structures, such as alpha (α) helices or beta (β) strands. The tertiary structure is formed by packing such structural elements into one or several compact globular units called domains. The final protein may contain several polypeptide chains arranged in a quaternary structure. By formation of such tertiary and quaternary structure amino acids far apart in the sequence are brought close together in three dimensions to form a functional region, an active site.

Protein folding remains a problem because there are 20 different amino acids that can be combined into many more different proteins than there are atoms in the known universe. In addition there is a vast number of ways in which similar structural domains can be generated in proteins by different amino acid sequences. By contrast, the structure of DNA, made up of only four different nucleotide building blocks that occur in two pairs, is relatively simple, regular, and predictable.

Since the three-dimensional structures of individual proteins cannot be predicted, they must instead be determined experimentally by x-ray crystallography, electron crystallography or nuclear magnetic resonance (NMR) techniques. Over the past 30 years the structures of more than 6000 proteins have been solved, and the sequences of more than 500,000 have been determined. This has generated a body of information from which a set of basic principles of protein structure has emerged. These principles make it easier for us to understand how protein structure is generated, to identify common structural themes, to relate structure to function, and to see fundamental relationships between different proteins. The science of protein structure is at the stage of taxonomy where we can begin to discern patterns and motifs among the relatively small number of proteins whose three-dimensional structure is known.

The first six chapters of this book deal with the basic principles of protein structure as we understand them today, and examples of the different major classes of protein structures are presented. Chapter 7 contains a brief discussion on DNA structures with emphasis on recognition by proteins of specific nucleotide sequences. The remaining chapters illustrate how during evolution different structural solutions have been selected to fulfill particular functions.

Proteins are polypeptide chains

All of the 20 amino acids have in common a central carbon atom (C_α) to which are attached a hydrogen atom, an amino group (NH_2), and a carboxyl group (COOH) (Figure 1.2a). What distinguishes one amino acid from another is the side chain attached to the C_α through its fourth valence. There are 20 different side chains specified by the genetic code; others occur, in rare cases, as the products of enzymatic modifications after translation.

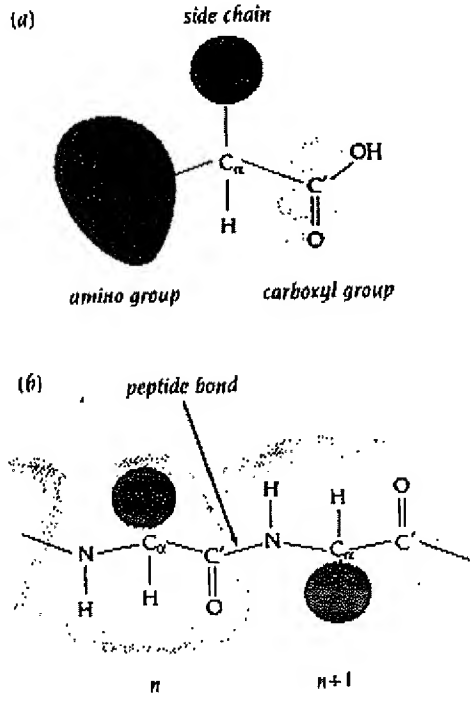
Amino acids are joined end-to-end during protein synthesis by the formation of peptide bonds when the carboxyl group of one amino acid condenses with the amino group of the next to eliminate water (Figure 1.2b). This process is repeated as the chain elongates. One consequence is that the amino group of the first amino acid of a polypeptide chain and the carboxyl group of the last amino acid remain intact, and the chain is said to extend from its amino terminus to its carboxy terminus. The formation of a succession of peptide bonds generates a "main chain," or "backbone," from which project the various side chains.

The main-chain atoms are a carbon atom C_α to which the side chain is attached, an NH group bound to C_α, and a carbonyl group C′=O, where the carbon atom C′ is attached to C_α. These units, or residues, are linked into a polypeptide by a peptide bond between the C′ atom of one residue and the nitrogen atom of the next (see Figure 1.2b). The basic repeating unit along the main chain from a biochemical or genetic viewpoint is thus (NH–C_αH–C′=O), which is the residue of the common parts of amino acids after peptide bonds have been formed (see Figure 1.2b).

The genetic code specifies 20 different amino acid side chains

The 20 different side chains that occur in proteins are shown in Panel 1.1 (pp 6–7). Their names are abbreviated with both a three-letter and a one-letter code, which are also given in the panel. The one-letter codes are worth memorizing, as they are widely used in the literature. A mnemonic device for linking the one-letter code to the names of the amino acids is given in Panel 1.1.

Figure 1.2 Proteins are built up by amino acids that are linked by peptide bonds to form a polypeptide chain. (a) Schematic diagram of an amino acid, illustrating the nomenclature used in this book. A central carbon atom (C_α) is attached to an amino group (NH_2), a carboxyl group (COOH), a hydrogen atom (H), and a side chain (R). (b) In a polypeptide chain the carboxyl group of amino acid *n* has formed a peptide bond, C–N, to the amino group of amino acid *n* + 1. One water molecule is eliminated in this process. The repeating units, which are called residues, are divided into main-chain atoms and side chains. The main-chain part, which is identical in all residues, contains a central C_α atom attached to an NH group, a C′=O group, and an H atom. The side chain R, which is different for different residues, is bound to the C_α atom.



The amino acids are usually divided into three different classes defined by the chemical nature of the side chain. The first class comprises those with strictly hydrophobic side chains: Ala (A), Val (V), Leu (L), Ile (I), Phe (F), Pro (P), and Met (M). The four charged residues, Asp (D), Glu (E), Lys (K), and Arg (R), form the second class. The third class comprises those with polar side chains: Ser (S), Thr (T), Cys (C), Asn (N), Gln (Q), His (H), Tyr (Y), and Trp (W). The amino acid glycine (G), which has only a hydrogen atom as a side chain and so is the simplest of the 20 amino acids, has special properties and is usually considered either to form a fourth class or to belong to the first class.

The four groups attached to the C_α atom are chemically different for all the amino acids except glycine, where two H atoms bind to C_α. All amino acids except glycine are thus chiral molecules that can exist in two different forms with different "hands," L- or D-form (Figure 1.3).

Biological systems depend on specific detailed recognition of molecules that distinguish between chiral forms. The translation machinery for protein synthesis has evolved to utilize only one of the chiral forms of amino acids, the L-form. All amino acids that occur in proteins therefore have the L-form. There is, however, no obvious reason why the L-form was chosen during evolution and not the D-form.

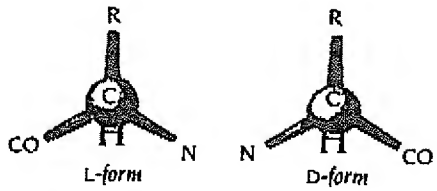


Figure 1.3 The "handedness" of amino acids. Looking down the H–C_α bond from the hydrogen atom, the L-form has CO, R, and N substituents from C_α going in a clockwise direction. There is a mnemonic to remember this; for the L-form the groups read CORN in clockwise direction.

Cysteines can form disulfide bridges

Two cysteine residues in different parts of the polypeptide chain but adjacent in the three-dimensional structure of a protein can be oxidized to form a **disulfide bridge** (Figure 1.4). The disulfide is usually the end product of air oxidation according to the following reaction scheme:

$$2\ \text{–}CH_2SH + \tfrac{1}{2}\ O_2 \rightleftharpoons \text{–}CH_2\text{–}S\text{–}S\text{–}CH_2 + H_2O$$

This reaction requires an oxidative environment, and such disulfide bridges are usually not found in intracellular proteins, which spend their lifetime in an essentially reductive environment. Disulfide bridges do, however, occur quite frequently among extracellular proteins that are secreted from cells, and in eucaryotes, formation of these bridges occurs within the lumen of the endoplasmic reticulum, the first compartment of the secretory pathway.

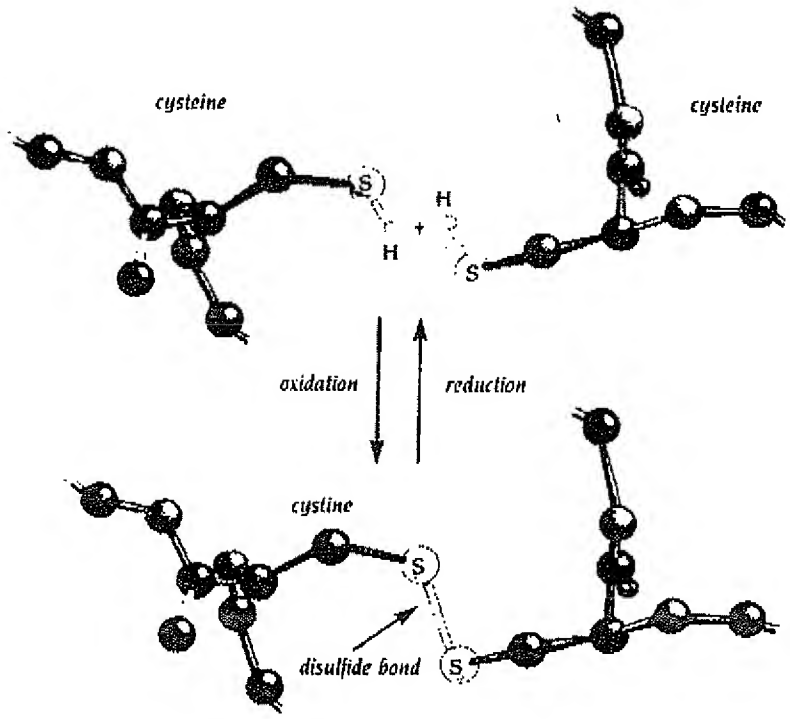
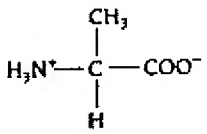


Figure 1.4 The disulfide is usually the end product of air oxidation according to the following schematic reaction scheme:

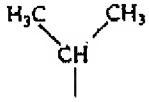
$$2\ \text{–}CH_2SH + \tfrac{1}{2}\ O_2 \rightleftharpoons \text{–}CH_2\text{–}S\text{–}S\text{–}CH_2 + H_2O$$

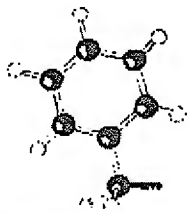
Disulfide bonds form between the side chains of two cysteine residues. Two SH groups from cysteine residues, which may be in different parts of the amino acid sequence but adjacent in the three-dimensional structure, are oxidized to form one S–S (disulfide) group.

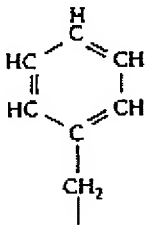


Ala, Alanine

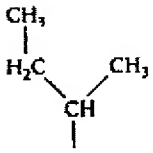
Val, Valine



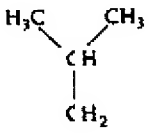




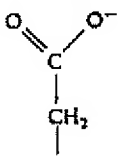
Phe, Phenylalanine



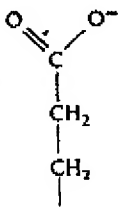
Ile, Isoleucine



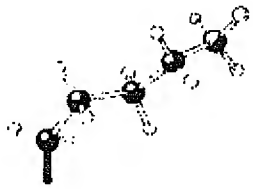
Leu, Leucine

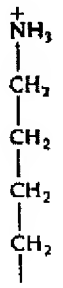


Asp, Aspartic acid



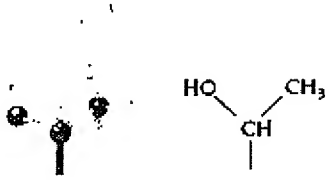
Glu, Glutamic acid



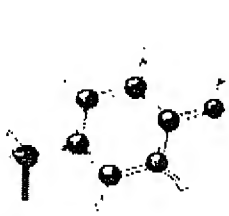


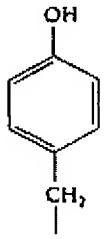
Lys, Lysine

Ser, Serine

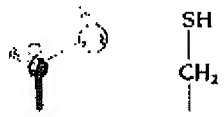


Thr, Threonine

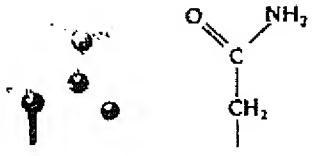




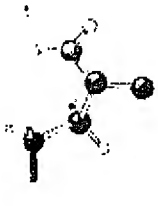
Tyr, Tyrosine

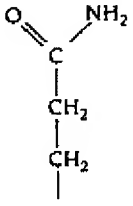


Cys, Cysteine



Asn, Asparagine





Gln, Glutamine

Folding and Flexibility

6

A protein, as we have seen, is a polypeptide chain folded into one or more domains, each of which is made up of α helices, β sheets and loops. The process by which a polypeptide chain acquires its correct three-dimensional structure to achieve the biologically active native state is called protein folding. Although some polypeptide chains spontaneously fold into the native state, others require the assistance of enzymes, for example, to catalyze the formation and exchange of disulfide bonds; and many require the assistance of a class of proteins called chaperones. A chaperone binds to a partly folded polypeptide chain and prevents it from making illicit associations with other folded or partly folded proteins, hence the name chaperone. A chaperone also promotes the folding of the polypeptide chain it holds. After a polypeptide has acquired most of its correct secondary structure, with the α-helices and β-sheets formed, it has a looser tertiary structure than the native state and is said to be in the molten globular state. The compaction that is necessary to go from the molten globular state to the final native state occurs spontaneously.

Protein folding generates a particular three-dimensional structure from an essentially linear, one-dimensional structure—a polypeptide chain with a particular sequence of amino acid residues. How to predict the three-dimensional structure of a protein from its amino acid sequence is the major unsolved problem in structural molecular biology. If we had a general solution to the protein folding problem, it would be possible to write a computer program to simulate protein folding and generate the precise three-dimensional structure of any protein from its amino acid sequence. However, a general solution to the folding problem is still not in sight, even though the number of proteins whose three-dimensional structure has been solved experimentally, in other words, the database of known protein structures, is doubling every 2 years.

A protein in its native state is not static. The secondary structural elements of the domains as well as the entire domains continually undergo small movements in space, either fluctuations of individual atoms or collective motions of groups of atoms. Furthermore, the functional activities of many proteins depend upon large conformational changes triggered by ligand binding. In this chapter, after discussing protein folding, we shall examine some examples of functionally important conformational changes of proteins.

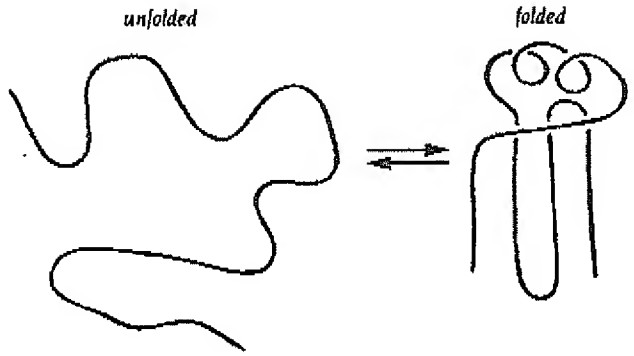


Figure 6.1 A polypeptide chain is extended and flexible in the unfolded, denatured state whereas it is globular and compact in the folded, native state.

Globular proteins are only marginally stable

Every biochemist or molecular biologist who has worked with proteins knows by experience that they are unstable. Slight changes in pH or temperature can convert a solution of biologically active protein molecules in their native state to a biologically inactive denatured state. The energy difference between these two states in physiological conditions is quite small, about 5–15 kcal/mol, not much more than the energy contribution of a single hydrogen bond, which is of the order of 2–5 kcal/mol.

There are two major contributors to the energy difference between the folded and the denatured state: enthalpy and entropy. Enthalpy derives from the energy of the noncovalent interactions within the polypeptide chain—the hydrophobic interactions, hydrogen bonds and ionic bonds. The covalent bonds within and between the amino acid residues in the polypeptide chain are the same in the native and denatured states, with the exceptions of disulfide bonds in those proteins where these form between cysteine residues. The noncovalent interactions on the other hand differ significantly between the two states. In the native state these interactions are maximized to produce a compact globular molecule with a tightly packed hydrophobic core whereas the denatured state is more open and the side chains are more loosely packed (Figure 6.1). These noncovalent interactions are therefore stronger and more frequent in the native state and hence their energy contribution, enthalpy, is much larger. The enthalpy difference between native and denatured states can reach several hundred kcal/mol.

Entropy derives from the second law of thermodynamics which states that energy is required to create order. Proteins in the native state are highly ordered in one main conformation whereas the denatured state is highly disordered, with the protein molecules in many different conformations. A typical experimental preparation of unfolded protein (a solution in 6 M guandinium chloride or 8 M urea) contains 10^{15}–10^{20} protein molecules, each of which will have a unique conformation. In the absence of compensating factors it would therefore be entropically much more favorable for the protein to be in the disordered denatured state. The energy difference due to entropy between the native ordered state and the denatured state can also reach several hundred kcal/mole but in the opposite direction to the enthalpy difference. The total energy difference between the native and the denatured state of 5–15 kcal/mol, which is called the free energy difference, is thus a difference between two large numbers, the enthalpy difference and the entropy difference. The fact that this difference is very small is a severe complicating factor both for predictions of possible native states and for interpretation of factors responsible for the stability or instability of protein molecules, because our knowledge about the denatured state is very incomplete.

We know much more about factors that influence the stability of the native state, mainly from experiments using directed mutations in proteins of known three-dimensional structure. Such experiments have yielded

precise information about energy contributions to the stability of the native state from close packing of hydrophobic side chains in the interior of the protein, and from the presence of disulfide bridges and interior hydrogen bonds and salt bridges, as well as from side chains that compensate the dipole moment of α helices (see Chapter 3).

The marginal stability of the native state over the denatured state is biologically very important. Living cells need globular proteins in correct quantities at appropriate times. It is therefore as important to be able easily to degrade these proteins as it is to be able to synthesize them. Globular proteins in living cells usually have a rather rapid turnover and their native states have therefore evolved to be only marginally stable. Moreover, the catalytic activities of enzymes, and other important functions of proteins, generally require some structural flexibility, which would be inconsistent with a rigidly stabilized structure.

Kinetic factors are important for folding

High resolution x-ray structure determinations of several hundred proteins have shown that in each case the specific sequence of a polypeptide chain appears to yield only a single, compact, biologically active fold in the native state. This fold generally has many substates with minor structural differences between them, as will be discussed later in this chapter, but all of these substates have the same general fold. Comparisons with structure determinations in solution by NMR show that the same fold also prevails in solution. In other words, under physiological conditions there appears to be one conformation for a given amino acid sequence that has a significantly lower free energy than any other. How is this folded state reached?

Intuitively one might imagine that all protein molecules search through all possible conformations in a random fashion until they are frozen at the lowest energy in the conformation of native state. The biophysicist Cyrus Levinthal showed in 1968 by a simple calculation that this is impossible. Assume as a gross simplification that each peptide group has only three possible conformations, the allowed regions α, β and l. in the Ramachandran diagram (see Figure 1.7), and that it converts one conformation into another in the shortest possible time, one picosecond (10^{-12} seconds). A polypeptide chain of 150 residues would then have $3^{150} = 10^{68}$ possible conformations. To search all these conformations would require 10^{48} years (10^{56} seconds)—an astronomical number compared with the actual folding time, which is between 0.1 and 1000 seconds both *in vivo* and *in vitro*. To occur on this short time scale, the folding process must be directed in some way through a kinetic pathway of unstable intermediates to escape sampling a large number of irrelevant conformations.

Such a folding mechanism raises several important questions that are difficult to examine experimentally, since the possible intermediates have a very short lifetime. If kinetic factors are important for the folding process it is possible that the observed folded conformation is not the one with the lowest free energy but rather the most stable of those conformations that are kinetically accessible. The protein might be kinetically trapped in a local low energy state with a high energy barrier that prevents it from reaching the global energy minimum which might have a different fold. In such a case structure prediction by energy calculations would give the wrong structure even if such calculations could be made with great accuracy. One important question therefore is how a living cell can prevent the folding pathway from becoming blocked at an intermediate stage. The most common obstacles to correct folding seem to be (1) aggregation of the intermediates through exposed hydrophobic groups, (2) formation of incorrect disulfide bonds, and (3) isomerization of proline residues. To circumvent these three obstacles cells produce special proteins that assist the folding process, as we shall discuss later in this chapter.

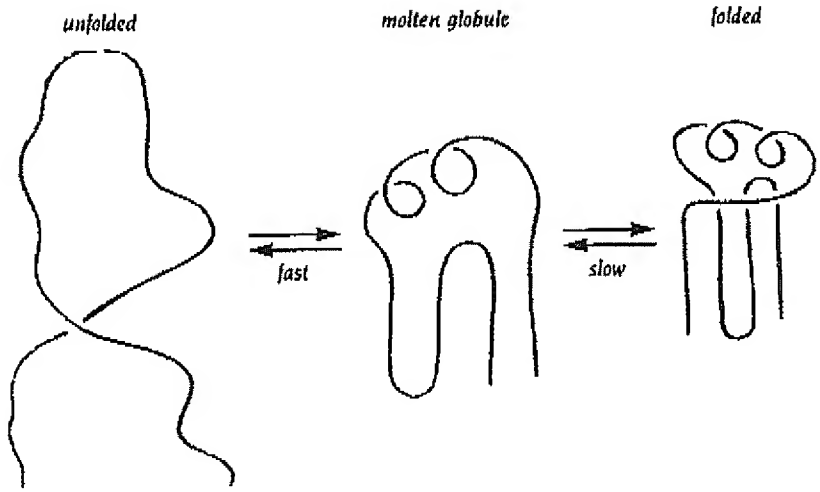


Figure 6.2 The molten globule state is an important intermediate in the folding pathway when a polypeptide chain converts from an unfolded to a folded state. The molten globule has most of the secondary structure of the native state but it is less compact and the proper packing interactions in the interior of the protein have not been formed.

An alternative way to remove kinetic barriers is exemplified by α-lytic protease, a bacterial enzyme which belongs to the serine protease superfamily of enzymes (Chapter 11). Like many other proteases it is synthesized and folded *in vivo* as an inactive precursor protein with a prosegment of 77 residues. This segment is excised after folding to produce the active enzyme. Unfolded precursor protein refolds easily *in vitro* but unfolded α-lytic protease lacking the prosegment does not refold. However, a solution of unfolded enzyme can be induced to refold by adding the excised prosegment. The capacity for folding obviously exists in the unfolded enzyme but there is a barrier present somewhere in the folding pathway that prevents folding. The prosegment removes this kinetic barrier, presumably by interacting with the enzyme in the unfolded state and thereby lowering the free energy of the transition states for folding; just as enzymes lower the free energy of transition states for chemical reactions and thereby increase the rates of the reactions (see Chapter 11).

Molten globules are intermediates in folding

The first observable event in the folding pathway of at least some proteins is a collapse of the flexible disordered unfolded polypeptide chain into a partly organized globular state, which is called the **molten globule** (Figure 6.2). This event is fast, usually within the deadtime of the experimental observation, which is a few milliseconds. We therefore know almost nothing about the process that leads to the molten globule, but we know some of the properties of this state. The molten globule has most of the secondary structure of the native state and in some cases even native-like positions of the α helices and β strands. It is less compact than the native structure and the proper packing interactions in the interior of the protein have not been formed. The interior side chains may be mobile, more closely resembling a liquid than the solid-like interior of the native state. Also loops and other elements of surface structure remain largely unfolded, with different conformations. The molten globule should, therefore, not be viewed as a single structural entity but as an ensemble of related structures that are rapidly interconverting (see Figure 6.3a).

In a second step, which can last up to 1 second, persistent native-like elements of tertiary structure begin to develop, possibly in the form of subdomains that are not yet properly docked. The ensemble of conformations is much reduced compared with those of the molten globule but it is still far from a single form. The single native form is reached in the final stage of folding, which involves the formation of native interactions throughout the protein, including hydrophobic packing in the interior as well as the fixation of surface loops.

Burying hydrophobic side chains is a key event

The collapse of the unfolded state to generate the molten globule embodies the main mystery of protein folding. What is the driving force behind the choice of native tertiary fold from a randomly oriented polypeptide chain?

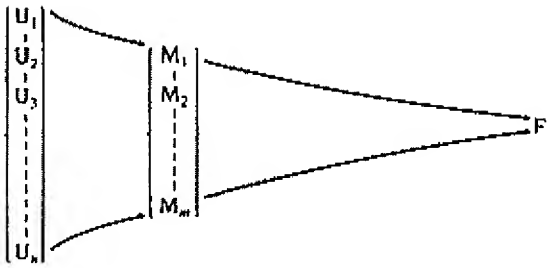
There is very little change in free energy by forming the internal hydrogen bonds that are characteristic of α helices and β sheets because in the unfolded state equally stable hydrogen bonds can be formed to water molecules. Secondary structure formation therefore cannot be the thermodynamic driving force of protein folding. On the other hand there is a large free energy change by bringing hydrophobic side chains out of contact with water and into contact with each other in the interior of a globular entity. Thus the most likely scenario is that the polypeptide chain begins to form a compact shape with hydrophobic side chains at least partially buried very early in the folding process. This scenario has several important consequences. It vastly reduces the number of possible conformations that need to be searched because only those that are sterically accessible within this shape can be sampled. Second, when some of the side chains are partly buried, their polar backbone -NH and -CO groups are also buried in a hydrophobic environment unable to form hydrogen bonds to water. This is energetically unfavorable unless they form hydrogen bonds to each other, which they can only do if they are close together. The simplest way to form such bonds is by forming elements of secondary structure: α helices and β sheets. The formation of secondary structure early in the folding process can therefore be regarded as a consequence of burying hydrophobic side chains and not as a driving force for the formation of the molten globule.

Looking at the amino acid sequence of a globular protein one finds that hydrophobic side chains are usually scattered along the entire sequence in a seemingly random manner. In the native state of the folded protein about half of these side chains are buried in the interior and the rest are scattered on the surface of the protein, surrounded by hydrophilic side chains. The buried hydrophobic side chains are not clustered in the sequence but are scattered along the entire polypeptide chain. What causes these residues to be selectively buried during the early and rapid formation of the molten globule? This question must be answered before one can solve the folding problem and be able to predict the fold of a protein from its amino acid sequence.

Both single and multiple folding pathways have been observed

In order to understand fully any folding pathway, all states of the pathway must be characterized both structurally and energetically. The simplified diagram in Figure 6.3 illustrates that during the folding process the protein proceeds from a high energy unfolded state to a low energy native state through metastable intermediate states with local low energy minima separated by unstable transition states of higher energy. The characterization of these states is not trivial and many different experimental techniques are employed, including NMR, hydrogen exchange, spectroscopy and thermochemistry.

Recently Alan Fersht, Cambridge University, has developed a protein engineering procedure for such studies. The technique is based on investigation of the effects on the energetics of folding of single-site mutations in a protein of known structure. For example, if minimal mutations such as Ala to Gly in the solvent-exposed face of an α helix, destabilize both an intermediate state and the native state, as well as the transition state between them, it is likely that the helix is already fully formed in the intermediate state. If on the other hand the mutations destabilize the native state but do not affect the energy of the intermediate or transition states at all, it is likely that the helix is not formed until after the transition state.



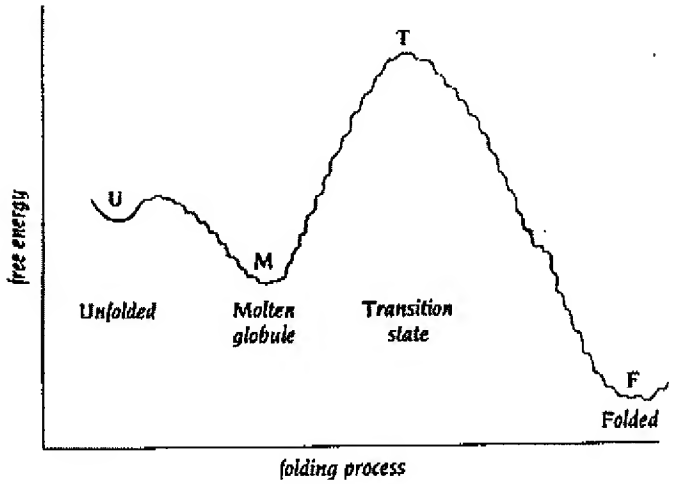


Figure 6.3 The unfolded state is an ensemble of a large number of conformationally different molecules, $U_1...U_n$, which undergo rapid interconversions. The molten globule is an ensemble of structurally related molecules, $M_1...M_m$, which are rapidly interconverting and which slowly change to a single unique conformation, the folded state F. During the folding process the protein proceeds from a high energy unfolded state to a low energy native state. The conversion from the molten globule state to the folded state is slow and passes through a high energy transition state, T.

The small bacterial ribonuclease, **barnase**, is a single chain protein with 110 amino acids and no disulfide bridges. Its three-dimensional structure was determined by the group of Guy Dodson, York University, and comprises three amino terminal α helices and a carboxy terminal five-stranded antiparallel β sheet (Figure 6.4). The group of Alan Fersht have examined the effects of mutations all along the structure and have made a detailed residue by residue characterization of its folding intermediate and transition states. They have concluded from their results that the intermediate molten globule state already has not only most of the native secondary structure elements but also the native-like relative positions of the α helix and β sheet as well as

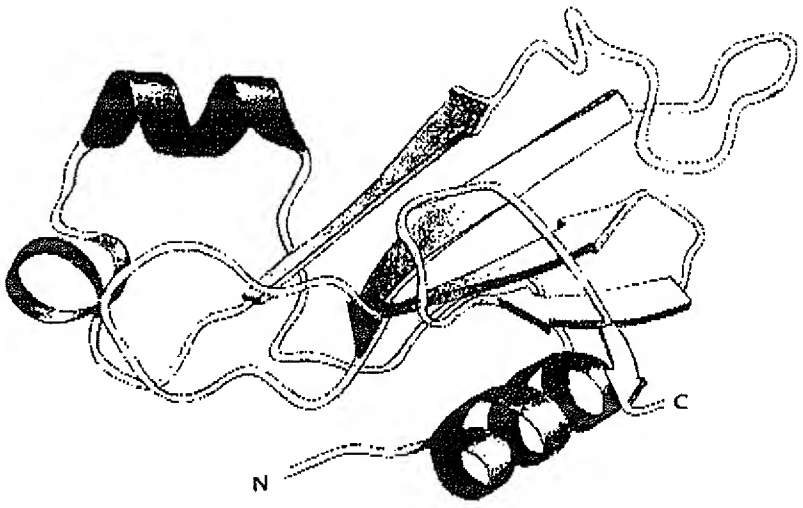


Figure 6.4 Schematic diagram of the structure of the enzyme barnase which is folded into a five stranded antiparallel β sheet (blue) and two α helices (red).

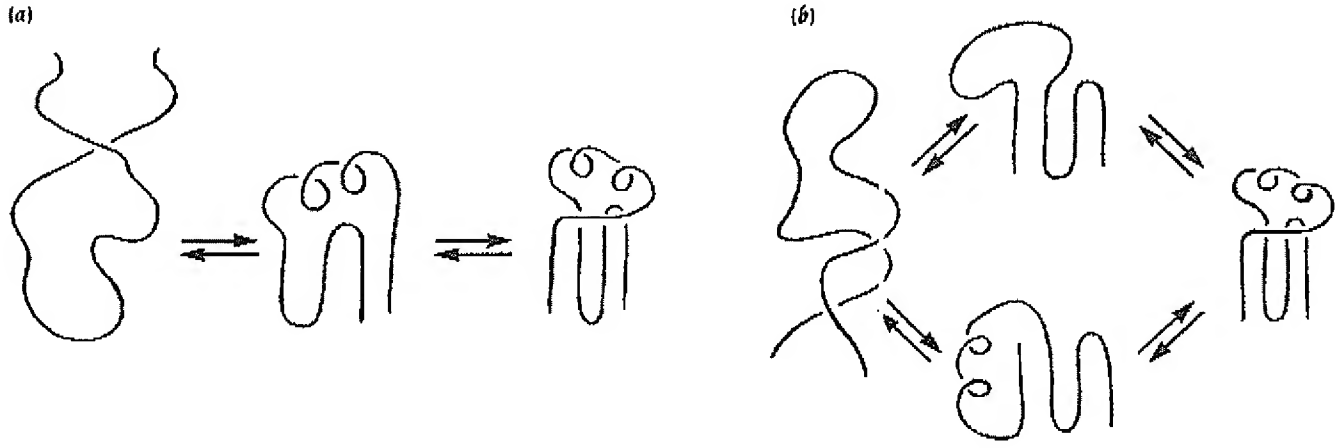


Figure 6.5 (a) Some proteins such as barnase fold through one major pathway whereas others fold through multiple pathways. (b) The folding of the enzyme lysozyme proceeds through at least two different pathways.

the relative positions of the β strands within the sheet. These results are consistent with the notion that the folding of barnase proceeds through a single major transition state and consequently through one major pathway (Figure 6.5a).

In contrast, folding of the enzyme lysozyme involves parallel pathways and distinct folding domains. Hen egg-white lysozyme was the first enzyme to have its structure determined crystallographically, in the laboratory of David Phillips then at the Royal Institution, London in 1965. The native structure consists of two lobes separated by a cleft (Figure 6.6). The first lobe comprises five α helices and the second is predominantly a three-stranded antiparallel β sheet. The folding of lysozyme has been studied extensively by a variety of complementary techniques (NMR, circular dichroism, fluorescence, hydrogen–deuterium exchange) to follow the development of different aspects of the structure such as formation of secondary structure, burial of hydrophobic aromatic groups and formation of hydrogen bonds. The group of Christopher Dobson, Oxford University, has used pulsed amide hydrogen–deuterium exchange to follow secondary structure formation. Amide hydrogen atoms are readily exchanged with the solvent in unfolded proteins, but this exchange is often strongly inhibited in a folded protein, especially for those amide groups that are hydrogen bonded in secondary structure elements. As a result, by measuring the rate of amide–hydrogen exchange as a function of folding time it is possible to monitor the formation of structure during the folding reaction. At 20 milliseconds, two major intermediate stages of lysozyme were detected: one in which the α-helical domain

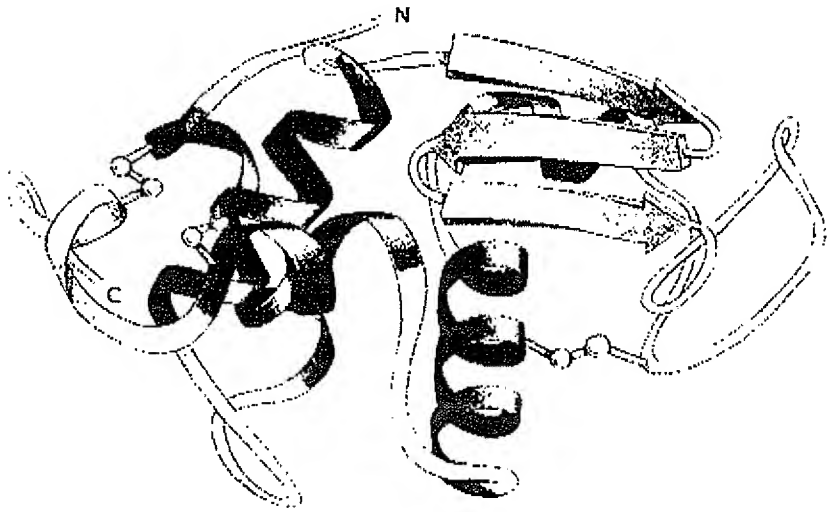


Figure 6.6 Schematic diagram of the structure of the enzyme lysozyme which folds into two domains. One domain is essentially α-helical whereas the second domain comprises a three stranded antiparallel β sheet and two α helices. There are three disulfide bonds (green), two in the α-helical domain and one in the second domain.